Examining the Equivalence of Simulated and Real AR on a Visual Following and Identification Task

Cha Lee,* Steffen Gauglitz,† Tobias Höllerer‡
Four Eyes Lab and University of California Santa Barbara
Doug A. Bowman§
Center for Human-Computer Interaction and Dept. of Computer Science, Virginia Tech

ABSTRACT

Mixed Reality (MR) simulation, in which a Virtual Reality (VR) system is used to simulate both the real and virtual components of an Augmented Reality (AR) system, has been proposed as a method for evaluating AR systems with greater levels of experimental control. However, factors such as the latency of the MR simulator may impact the validity of experimental results obtained with MR simulation. We present a study evaluating the effects of simulator latency on the equivalence of results from an MR simulator and a real AR system. We designed an AR experiment which required the participants to visually follow a virtual pipe around a small room filled with real targets and to find and identify the targets which were intersected by the pipe. We show that, with a 95% confidence interval, the results from all three simulated AR conditions fall well within one standard deviation of the real AR case.

Index Terms: H.5.2 [Information Systems]: User Interfaces—Evaluation/methodology; I.3.6 [Computing Methodologies]: Computer Graphics—Methodology and Techniques - Device independence

1 INTRODUCTION

MR simulation encapsulates the idea of using a high-fidelity Virtual Reality (VR) display system to simulate a range of lower fidelity display systems. To simulate different display systems (including other VR and Augmented Reality (AR) systems), researchers replicate the immersion level of the desired system. This concept has been used in the VR community [2, 1] to conduct research in quantifying the benefits of immersion for VR and AR systems. MR simulation is a good choice for this particular type of research because it affords three advantages: experimental control, experimental repeatability, and practical cost. Experimental control is achieved since the same actual display system is used to simulate all conditions. The variables of interest, such as field of view or stereoscopy, can be isolated while all other display characteristics are held constant. Experimental repeatability is achieved since other researchers can easily replicate the display systems used in other experiments with a single higher-fidelity VR system. Our position and hypothesis is that there are many situations in which MR simulation constitutes a valid approach, making it a fruitful and exact tool for AR usability evaluations. However, since these experiments are performed with simulated displays that may have subtle differences such as added latency, the validity of their results may be questioned. To that end, this paper presents an experiment in which we evaluate the effect of the simulator's end-to-end latency

---

*chalee21@cs.ucsb.edu

†sgauglitz@cs.ucsb.edu

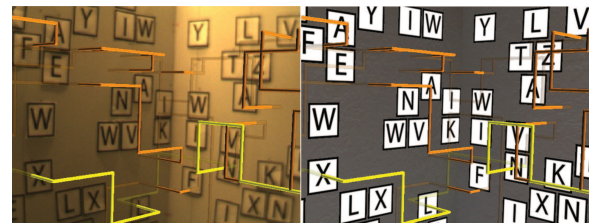‡holl@cs.ucsb.edu

§dbowman@vt.edu

Figure 1: Environment used in experiment. The left image is representative of the real AR condition. The right image is representative of the three simulated AR conditions.

on results from an AR experiment with a visual following and identification task.

2 EXPERIMENTAL DESIGN

In previous work, we found that simulator latency had a significant effect on user performance in an AR experiment that involved a 3D tracing task. Even when the visual registration between simulated virtual and simulated real world objects was the same, the mismatch between the simulator latencies (causing proprioceptive misregistration) was sufficient to affect results. These results indicate that in those cases, care must be taken to minimize the effects of simulator latency and to take the effects of simulator latency into consideration when analyzing and reporting results. While it is useful to know how simulator latency affects object manipulation, it is equally important to find tasks that are not affected by simulator latency. After careful consideration, we determined that a purely visual task in AR, particularly one concerned with following AR annotations and identifying physical objects they annotated, would be a good candidate. We designed an experiment where we placed the user in a real AR system and asked them to perform the task. We then placed the user in a virtual simulation of the same AR system, and asked them to perform the visual task while we introduced different simulator latencies.

Environment and Task   The hardware used for the experiment was an NVis SX111 wide-FOV HMD for display, InterSense IS900 for tracking, and WorldViz's Vizard development platform for rendering. The environment was a simple room with four virtual pipes (as seen in Figure 1). The room was approximately 3.4x2.9 m and had many targets taped to the walls. These targets were 19x19 cm sized papers printed with a large black letter and a bold, black outline to clarify where the target borders were. Four continuous pipes snaked around the walls, going through the walls and into the space between walls at numerous points. Three were orange distractors and one was yellow for the participant to follow. As the yellow pipe snaked its way in and out of the walls, it intersected with numerous targets. In the simulated AR conditions, users were situated in the same room but saw a virtual representation of the room instead. In the AR case, the overlay of the (virtual) pipes on the (real) targets was affected by hardware latency and calibration imperfec-
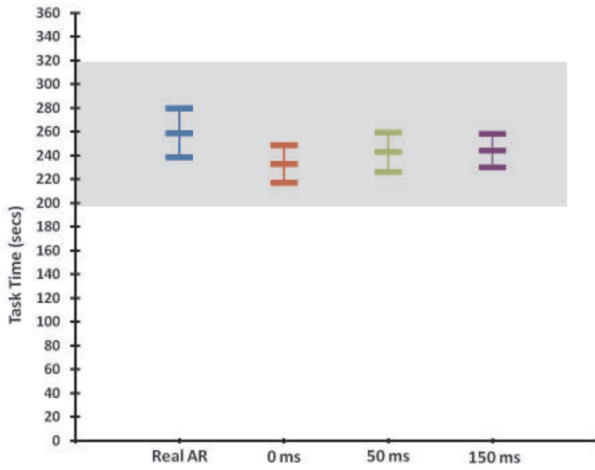
Figure 2: Means and 95% confidence intervals of the four experimental conditions. The gray area represents represents a zone of scientific indifference which we set at one standard deviation from the real AR condition task time.

tion, but by using a highly accurate computer-vision based tracker we verified that the intersection point never "wanders off" the target despite registration errors under extreme viewing angles.

Through a series of pilot studies, we created a customized X-ray vision following and identification task. The yellow pipe always began at the same letter marker, and this beginning point was designated by a green ball attached to the end of the pipe. Before the start of the task, all pipes would be hidden from the user, including the distractor pipes. Once the user indicated they were ready, the study administrator would start the task by showing the pipes, which also started the timer. The participant would then follow the yellow pipe and if the pipe intersected with a target letter, the participant would verbally indicate the letter. This would proceed until the participant reached the end of the yellow pipe, which was designated by a red ball. Once there, the participant would say "done" and the study administrator would stop the timer. If a mistake was made at any time during the task (missing a letter or saying the wrong letter), the administrator would inform the participant of the first error and where it had occurred. This pipe was then placed at the back of the pipe queue, to be done again during the same session of trials.

Participants and Procedures We had a total of 26 participants. The participants were evenly split in terms of gender, with 13 males and females each. The average age of all the participants was 21.19 years old, and all were between the ages of 18 and 39. All participants reported they had correct or corrected vision. A pre-questionnaire was administered prior to the study. From this demographic information, we can say that our users were mostly unfamiliar with AR and HMDs, but had a passing familiarity with 3D video games and VR. Participants were compensated for their time by a nominal amount of $10 per hour.

In the experiment, we conducted a training session (with the 0 ms simulated latency condition) followed by timed sessions with breaks in between. The timed sessions consisted of four sessions, one for each condition. During each session, the condition of the study was kept constant (real AR, simulated AR with 0, 50, or 150 ms) and the participants were asked to complete six pipes. Once all six pipes had been done correctly with zero mistakes, a break was given.

3 RESULTS

Task Time As seen in Figure 2, the average task time for all four scenarios were very similar. All four scenarios had results

which were well in the range of one standard deviation from the real AR task time. The real AR conditions took participants slightly longer to complete, with the 50 ms and 150 ms conditions taking approximately the same amount of time, and the 0 ms condition having the shortest task completion time.

Based on the literature [3], we determined that an equivalence test using confidence intervals was the best method for this analysis. This method for determining equivalence also requires a known equivalence threshold, which can only be obtained through experience and previous research. As there is no such threshold for a task and experiment of this nature, we decided to set this threshold by using the standard deviation of the data from the real AR condition. The real AR condition represents the true experiment and in a sense is the ground truth for the results of the other three scenarios. By plotting the confidence intervals of the four scenarios against this zone of scientific indifference, there is evidence that the conditions may be scientifically equivalent. However our sample size is small when compared to clinical studies which test clinical equivalence. But even with our 26 users, already we do obtain very consistent results. By observing Figure 2, the confidence intervals of each condition overlap and are relatively compact. This is a strong indication that our equivalence results may hold.

Scenarios	Mean Time	Stdev Time	Sum Errors
Real AR	258.90 s	60.82 s	43
0 ms	232.90 s	46.83 s	25
50 ms	242.70 s	49.85 s	22
150 ms	243.80 s	42.04 s	35

Table 1: Mean task time, standard deviation, and total number of errors for each condition. Each error represents a re-do for a pipe in that condition.

Error Rate Table 1 shows the total number of errors for each condition. As can be seen, the real AR condition accrued the most errors at 43. These errors may be indications of the qualitative differences between the scenarios and of learning effects which may have occurred since all users were trained in the simulated AR condition.

4 CONCLUSIONS AND FUTURE WORK

We have presented a user experiment evaluating the effect of simulator latency on a path following task in simulated AR. Our initial results show that at a threshold of one standard deviation from the true AR condition, the simulated AR conditions were equivalent to the true AR condition. In future work, we plan to re-do this experiment with more users using a between subjects design. We also plan on looking into the effect of visual realism on simulated AR experiments.

REFERENCES

[1] D. A. Bowman and R. P. McMahan. Virtual reality: How much immersion is enough? *Computer*, 40(7):36–43, 2007.
[2] R. Pausch, D. Proffitt, and G. Williams. Quantifying immersion in virtual reality. In *SIGGRAPH '97: Proceedings of the 24th annual conference on Computer graphics and interactive techniques*, pages 13–18, New York, NY, USA, 1997.
[3] G. Tunes Da Silva, B. R. Logan, and J. P. Klein. Methods for equivalence and noninferiority testing. *Biology of blood and marrow transplantation journal of the American Society for Blood and Marrow Transplantation*, 15(1 Suppl):120–127, 2009.